Enriching machine translation input using semantics-based fuzzy matches

Tom Vanallemeersch and Leen Sevens

Computer-aided translation tools allow translators to look up source sentences in a translation memory (TM) and retrieve similar source sentences (fuzzy matches) and their translation. Fuzzy matching can be integrated with machine translation (MT) by pretranslating the matched source parts in order to have the MT system focus on the non-matched parts. Alignment links between a source sentence in the TM and its translation allow for this kind of pretranslation. We compare the MT results produced using the basic type of fuzzy matching, Levenshtein distance, with the results produced using two types of semantics-based fuzzy matching, in the context of English to Dutch translation and vice versa. The first of these types consists of matching based on semantic predicates and roles, while the second type applies lexical semantic information, more specifically WordNet and paraphrase tables, to compare sentences based on the relatedness of words and word groups. We perform the second type of matching through the METEOR metric, which was originally designed for evaluating MT systems.